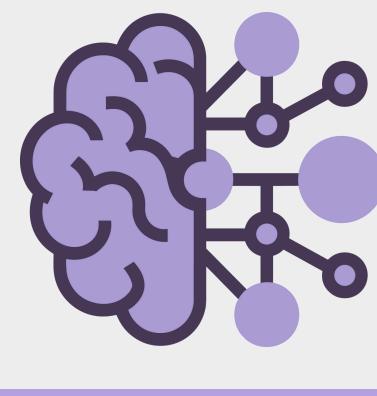


REDES NEURONALES ARQUITECTURAS

Análisis Semejanzas y Diferencias



DEFINICIÓN

Una red neuronal artificial (RNA) es un modelo computacional inspirado en el cerebro humano, compuesto por nodos (neuronas artificiales) interconectados en capas que procesan información.

ARQUITECTURA DENSA

Se trata de una serie de capas, en donde cada una está conformada por Perceptrones, neuronas artificiales que tienen un peso y una suma ponderada y que están conectadas entre sí de forma secuencial, aprendiendo de la capa anterior.

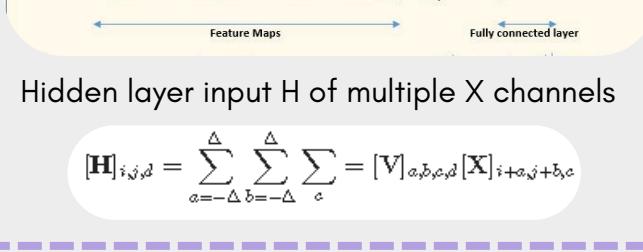
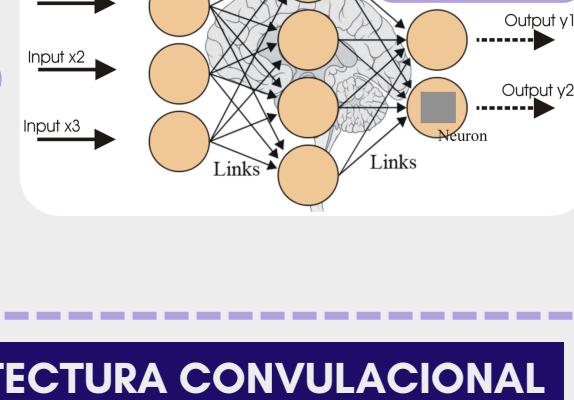
Clasificación Regresión Redes pequeñas Problemas tabulares

Modelo lineal Perceptrón

$$\hat{y} = \text{sign}(w^T x)$$

Gradiente Descendiente

$$w^* = \arg \max_w - \sum_{x_n \in X} \langle w^T x_n, y_n \rangle$$



Hidden layer input H of multiple X channels

$$[H]_{i,j,d} = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} \sum_{c} = [V]_{a,b,c,d} [X]_{i+\alpha, j+\beta, c}$$

ARQUITECTURA CONVULACIONAL

Es una extensión de la arquitectura densa, pero por cada capa se divide la estructura (espacial) en regiones locales y se realiza una convolución entre las características. Las primeras capas se fijan en detalles sencillos y las últimas en patrones y objetos concretos.

Visión por computadora y satelital

Ánalisis video y audio

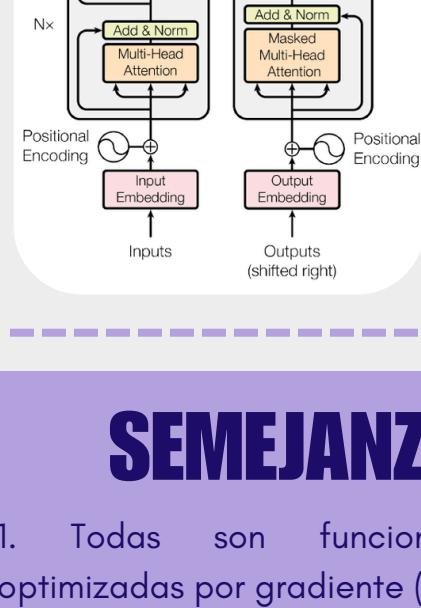
ARQUITECTURA RECURRENTE

Las RNN funcionan reteniendo información de entradas anteriores en un "bucle de retroalimentación" para que cada paso en la secuencia influya en la salida. Se trabaja en procesos que tienen relación con series de tiempo y presentan dos tipos:

Autorregresivos

Autorregresivos de Variable Latente

$$x_t \sim P(x_t | x_{t-1}, \dots, x_{t-\tau}) \quad \hat{x}_t = P(x_t | h_t) \quad h_t = g(h_{t-1}, x_{t-1})$$



ARQUITECTURA BASADAS EN ATENCIÓN

La atención es una operación que permite que un modelo aprenda qué partes de una secuencia son relevantes en cada momento. En lugar de procesar la información en orden (como una RNN), la atención permite mirar todas las posiciones simultáneamente y asignarles diferentes pesos de importancia.

Procesamiento de Lenguaje Natural

Video y audio

Robótica y control

Scale dot formula Attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

SEMEJANZAS

- Todas son funciones parámetricas optimizadas por gradiente (backpropagation).
- Todas usan no linealidades (funciones de activación). Ejemplos: ReLU, tanh, sigmoid, GELU.
- Todas trabajan con representaciones vectorizadas, aunque cambie la forma del input.
- Todas pueden funcionar como aproximadores universales.
- Todas pueden ampliarse en profundidad y anchura.

DIFERENCIAS

- No todas entienden el mismo contexto (estructuras espaciales y temporales y dependencias globales o locales).
- El procesamiento es distinto para cada uno, ya sea que es secuencial, convolucional o análisis en paralelo (como los transformers).
- El número de parámetros es distinto, ya sea de los pesos de los perceptrones, de los filtros y de los pasos temporales.
- El costo computacional y la parallelización varían según cada arquitectura.
- Cada arquitectura tiene un dominio en especial en donde destaca (tablas, imágenes, audio... etc).